DOSSIER

DATA-DRIVEN JOURNALISTIC OPERATION:

reshaping the idea of news values with algorithms, artificial intelligence and increased personalization

MÁRCIO CARNEIRO DOS SANTOS
Universidade Federal do Maranhão, São Luís – Maranhão – Brazil
ORCID: 0000-0002-0663-1300

DOI: 10.25200/BJR.v16n3.2020.1295

Received on: April 03rd, 2020
Desk Reviewed: May 04th, 2020
Desk Review Editors: Laura Storch and David Domingo
Revised on: August 17th, 2020
Approved on: November 03rd, 2020

ABSTRACT – This article discusses the strategic reconfiguration of data in news organizations through the use of algorithms and artificial intelligence, more specifically machine learning, to increase the idea of value around information, which has suffered from fragmented audiences, a massive increase in the number of broadcasters, indirect competition from large technology companies, and changes to the digital media ecosystem. We propose to identify patterns of interest, predict social engagement, and allocate resources for new coverage as forms for increasing the current level of personalization offered to news consumers.
Key words: algorithms, artificial intelligence, machine learning

A OPERAÇÃO JORNALÍSTICA GUIADA POR DADOS: reconfigurando a percepção de valor da notícia com algoritmos, inteligência artificial e personalização estendida

RESUMO – Discute-se a reconfiguração estratégica da utilização de dados dentro das organizações jornalísticas, através de algoritmos e soluções de inteligência artificial, entre elas especificamente o aprendizado de máquina, como alternativa para elevação da percepção de valor sobre o produto informativo, reduzida pela fragmentação das audiências, explosão de emissores, concorrência indireta das grandes empresas de tecnologia e transformações do ecossistema de meios digitais. Tal abordagem propõe a identificação de padrões de

 DOI: 10.25200/BJR.v16n3.2020.1295

interesse, a predição de engajamento social e a alocação de recursos para coberturas como formas de expandir o atual nível de personalização oferecido aos consumidores de notícia.
Palavras-chave: Algoritmos. Inteligência artificial. Aprendizado de máquina.

LA OPERACIÓN PERIODÍSTICA GUIADA POR DATOS: reconfigurando una percepción del valor de la notificación con algoritmos, inteligência artificial y personalização estendida

RESUMEN – Se discute la reconfiguración estratégica del uso de datos en las organizaciones periodísticas a través de algoritmos y soluciones de inteligencia artificial, específicamente, el aprendizaje automático como una alternativa para aumentar la percepción de valor sobre el producto de información, el cual se reduce por la fragmentación de las audiencias, el aumento en la cantidad de emisores, la competencia indirecta de grandes compañías tecnológicas y las transformaciones del ecosistema de medios digitales. Tal enfoque propone la identificación de patrones de interés, la predicción del compromiso social y la asignación de recursos para la cobertura como formas de expandir el nivel actual de personalización ofrecido a los consumidores de noticias.
Palabra clave: Algoritmos. Inteligencia artificial. Aprendizaje automático

1 Introduction

Over the last few decades a lot of academic material has been produced on the use of computer resources, at different levels, for producing news content. Computer-assisted reporting (CAR), database journalism (DBJ), data-driven journalism (DDJ), computer journalism and even robotic journalism are some of these resources.

The first three of these resources are connected to the continuous growth in the volume of information to which journalists are exposed to, with the excess amount of data available online and the impact it has on the construction of news.

The most utilitarian idea of the computer, as a tool used by journalists to help them throughout the various stages of their work, is what CAR is mainly used for.

There is also an abundance of literature on the use of databases and how it supports newsroom backends (the internal operations

behind distribution such as content management systems, publishing tools, and newsroom controls) which has already been consolidated in theoretical terms in DBJ (Barbosa, 2007, 2008, 2009, 2011; Fidalgo, 2004, 2007; Machado, 2006), citing the databases "as defining the structure and organization, as well as the composition and presentation of content of a journalistic nature" (Barbosa & Torres, 2013, p.154).

The concept of open data[1] (the transposition and consolidation of huge public databases for transparency and other similar portals), in addition to changes to national laws on the access to information and control over public spending, have reinforced the idea that journalists see digital repositories as sources that can be interviewed and as places to identify new stories which this data could help to tell. Much of what is called data-driven journalism today is based on these premises.

Closer to the concept of computer journalism, there are studies on the automation of the production process, describing how software is capable of writing journalistic texts, which have been conducted by Carreira (2017) who speaks about journalism made by non-humans, and Višnovský et al. (2019) who propose their idea of robojournalism, the interconnection between information technology, statistics, and reporting. There are further studies which use terms such as automated journalism and robotic journalism (Maier, 2002; Dalen, 2012; Carlson, 2014; Clerwall, 2014; Lewis & Usher, 2014; Latar, 2015; Santos, 2014; Santos, 2016a; Graefe, 2016) and deal with similar processes.

The internal perception about changes in the production process through the use of algorithms and automation is another aspect which has also been studied, found in works such as those by Rubio and Ruiz (2020) and López et al. (2018).

Coddington (2014) lists a few differences between these concepts:

a) Computer-aided reporting (CAR) is the oldest of these concepts. It is based on Meyer's ideas of precision journalism (1973), which is founded on his proposal of an activity developed on academic research techniques;

b) Data-driven journalism (DDJ) is the interaction between computer resources and the quantitative approach, which is used to write agendas and help journalists who are trained to work with spreadsheets, databases, visualizations, and even programming.

c) Computer journalism (CJ) deals with the development of algorithms, the automation of stages of the production process, and the eventuality of machines replacing journalists.

DOI: 10.25200/BJR.v16n3.2020.1295

Diakoupolous (2011) writes:

> Computational Journalism is the application of computing and computational thinking to the activities of journalism including information gathering, organization and sensemaking, communication and presentation, and dissemination and public response to news information". (Diakoupolous, 2011, p.1).

All these approaches, however, do not include or suggest another way to use the data and algorithms and perhaps connect them with the operations of a news organization as a whole.

This idea, which is the intention of this text, goes beyond news production; it focuses on the potential of using computer tools in the relationships that a news organization establishes with its external environment, these include all users, readers, consumers, subscribers, advertisers, agencies and the market in general. It also includes business models and the perception of value about the product the news organization delivers, outlining an approach which is not necessarily linked to just news content but to the economy of communication.

This perspective falls within the field of communication in general, as well as the fields of journalism, marketing, data science, and emerging technologies, as it is clearly interdisciplinary and complex.

Our goal is to analyze (in an exploratory way) the transition from emergent data and algorithms used by news organizations to the current level used by other types of industries, where data and algorithms have become a strategic factor of fundamental centrality not only for conquering or maintaining markets, but also for their own survival.

In this text we make a general theoretical proposition and discuss three possible results for reshaping an input type which has always been plentiful in newsrooms but historically underused: data. This is done through the implementation of an internal culture focused on generating value from information, not only in the production of news (which has always existed), but in several aspects of the journalism industry, including its operational routines, its relationship with consumers, the design of its products, and its forms of distribution.

2 The news industry and its crisis

In the analogue world, the business model of traditional print journalism was based on three sources: advertising, newsstand sales and subscription marketing. The electronic journalism of radio and TV grew from the delivery of apparently free content interspersed with commercial breaks between news segments which advertisers pay for indirectly in order to speak to the public.

Decades of journalistic activity were built on this solid foundation because businesses were generating decent revenue; they were producing enough resources to cover the costs of running the information production process while also generating financial and political profits through the significant influence they gained from delivering mass amounts of content with little to no competition.

The large communication groups in Brazil eventually grew out of this central situation where controlling the channels for disseminating information meant profit and power.

With the arrival of the internet, the news industry thought that everything would remain the same, or perhaps be even better. The new medium would be just another channel for delivering information and the established model would continue to operate with the same efficiency, reaching new audiences, generating even more revenue, and helping to expand the strength of the organizations that already controlled traditional channels.

Time showed us that this is not exactly how things would go. The transition to digital changed the paradigm of information consumption from one which was based on scarcity, centrality, homogeneity and unidirectionality to a very different one where the abundance of content, the decentralization of sources resulting from an explosion of 'broadcasters', the fragmentation of audiences and the possibilities of interaction and interference in the flow are all identifiable characteristics of the current stage of an accelerated and still incomplete transformation process.

Even before the onset of the Internet, large media outlets were huge hubs for information and attention, controlling unidirectional flows of information from the sender to its audience, and the audience gave its total attention to the content it received.

In the digital environment, however, the structure of the networks that connect people and media outlets has changed; they are now forced to deal with two-way interactions, where information and

  DOI: 10.25200/BJR.v16n3.2020.1295

audiences' attention are now performed equally in a complex system where the hegemony of the mainstream media has been reduced.

Information started to flow more intensely from users to media outlets through users' recent ability to become broadcasters themselves. Users are now capable of producing text and audiovisual content (most of the time from a cell phone) and have become not just sources but also collaborators, evaluators, partners in disseminating and sharing content, and indirect generators of metrics and insights about preferences that are so essential in today's newsrooms.

This context also meant that media attention would now be directed more intensely towards information consumers. That's not to say that media outlets were not already intensely focused on information consumers, but now, in addition to simply ensuring that their content reaches audience, they also have to fully understand those audiences and reach them in real time, using not only traditional contact channels but also social networks, video platforms, and any other digital pathway where data can be extracted.

The impact of these changes on the news industry has been devastating, and the ability of news organizations to adapt to this new scenario is constantly being tested.

Some newsrooms had to be closed down, others suffered drastic cutbacks. The discussion about a crisis in journalism was resumed on a number of aspects: ethics, technique, and in our view, perhaps the most harmful: the crisis of a business model that was proving to be inefficient in the face of the massive changes to the environment in which news organizations operate.

The application of emerging technologies such as Artificial Intelligence – AI (Coppin, 2010), the internet of things – IoT (Santos, 2016b; Santos, 2016c), virtual reality (Santos, 2019a), augmented reality (Santos, 2015), and especially big data (Santos, 2019b) is still in its early stages in news organizations, which is different from other industries.

We defend the hypothesis that the crux of the crisis in the journalism industry lies in the drastic reduction in the consumers' perception of value about news products. The past environment of scarcity that meant we had to leave home and go to a newsstand to buy the daily newspaper or that kept us in front of the TV at specific times of the day to get our fair share of information was slowly disappearing. The abundance of sources and informative content widely available via the internet, social networks, and a number of

other applications easily accessible by cell phones has made the need for traditional news products become almost obsolete.

The natural reaction of the industry was to face the fact that audiences were fragmented and revenues and centrality were being lost and review their operation models and processes. Some of these attempted updates had little or more success depending on the region, the size of the business, and the level of credibility and ability of the particular outlet to realize the task.

Switching subscriber bases from physical to digital, paywalls, connecting to social media platforms, updating websites, and other attempts at innovation such as producing immersive content are just some of the various ways in which news organizations tried to re-establish themselves in the face of a hostile scenario.

The fundamental search for information was now being offered by several other mechanisms, even though these mechanisms introduced disinformation, fake news, and a whole set of new problems into the media ecosystem.

Even though it is impossible to discuss the issue of deliberately producing disinformation (due to the limitations of this text) we would like to point out that the economic, social and technological transformations in the digital environment have made that environment the target of various agendas aimed at influencing, guiding or conducting public debates. This is done through the use of automation tools and an understanding of the internal logic of algorithms that control what people see on social media platforms, even to the point that it influences their results. Some aspects of this problem and how it affects journalism can be found at Nechushtai and Lewis (2019), Sánchez and Ruiz (2020) and Saquete et al. (2020).

The production of fake news, or disinformation, is not necessarily a new tactic in the history of mankind, but the contemporary information ecosystem has enhanced its impact potential.

The recent growth of initiatives and organizations for checking facts and combating false news, in addition to being fundamental, has also opened up a path to defend the need for traditional journalism and its process of mediation. Global crises such as the coronavirus pandemic have allowed traditional news products to return, at least partially, to the centrality and reach they used to have.

Given the general framework presented so far and considering the initial proposal that goes beyond just the production of news content, even if exploratory in nature, what

  DOI: 10.25200/BJR.v16n3.2020.1295

other actions could be taken in order to achieve better results from the newspaper industry and increase the relevance of their product in their consumers' eyes?

3 Data and algorithms and reshaping contact with users

The emerging technologies in the journalism industry mentioned above (which other industries place high strategic value on for their operations) are still mostly used for processing information that is to be converted into news content and not used for generating information about users' daily relationships with this product.

There are a number of possibilities (listed below) which are still underused for producing information content and represent a potentially more personalized and effective approach with users based on data from their interactions with publications.

Within the large field of artificial intelligence, we find segments such as machine learning (ML), computer vision, and natural language processing studies, just to name a few.

Grus (2016, p.142) defines the term machine learning as "referring to the creation and use of models that are learned from data". Thus, ML models use pre-existing data to try and "predict possible outputs for new data".

For example, this approach is commonly used to identify whether the emails we receive are spam or not. Amaral (2016, p.7) defines "classification, regression, groupings and rules of association" as some of the main tasks related to machine learning.

When classifying a new or unidentified element the objective is to use a pre-defined set of elements that share common attributes (which we call classes) to define which class this new element belongs to. According to Amaral (2016, p.7) "we use classification to predict fraud, to find out what species an animal belongs to, to predict a disease, or to classify a type of fungus".

Instead of working with categories or nominal data, regression works with numerical data and can be used to estimate a person's weight from the number of calories that person consumes on a daily basis.

There are no pre-existing classes for grouping or clustering, however, and we analyze the attributes of several elements in order

to identify groups or clusters of elements that share similar attributes and then divide them into groups.

> Depending on the type of grouping we use, an element may belong to more than one group or not be grouped, that is, be considered noise. Examples of use: identify groups of customers and target specific marketing campaigns, identify attempts to access the network; categorize a new species among others. (Amaral, 2016, p. 8).

Finally, we have the rules of association which try to identify relationships between items. A very common example is the practice in e-commerce of associating purchases between users. This is based on data showing that whoever bought item A also bought items B and C. This information is then used to suggest new items to consumers.

A number of types of algorithms are associated with ML models for its various applications. Bayes, Frequency of Subsets and Data Tree are some of these models (Amaral, 2016; Grus, 2016) and their details go beyond the scope of this text.

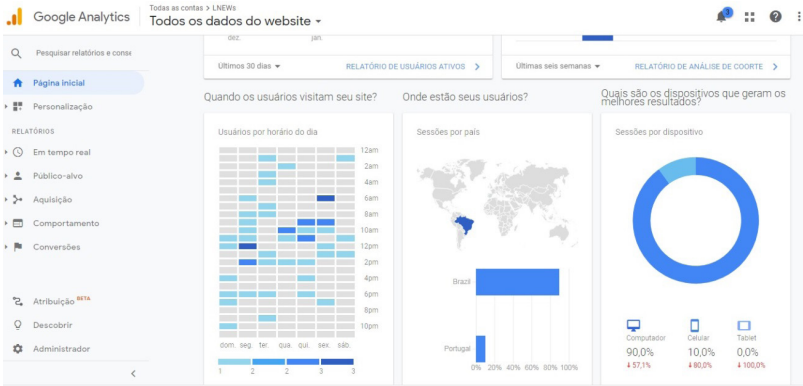
Most of the applications we discuss below are related to machine learning and how it has reshaped the relationship between newspaper companies and their users. We chose Google Analytics because it is free and the most widely adopted solution for generating metrics on the web. NewsWhip on the other hand is a startup that integrates the type of predictive analysis that we describe here in its solutions. It also has an extension for the Chrome browser that can be installed for testing by researchers at no cost.

3.1 Identifying patterns of interest

Pattern identification is one of the most common ways artificial intelligence is used. Nowadays with tools such as Google Analytics we can collect a diverse amount of data on users' behaviors when they interact with news products. This data includes the pages you visit, the time you spend on each page, the source of the connection, the type of device you use to access the content, the operating system, and even your geographic location.

DOI: 10.25200/BJR.v16n3.2020.1295

Figure 1 – Google Analytics data screen



Source: author's elaboration

Through ML tools this data set can be used to propose content, associate products based on consumption profiles, and notify users of any news in the categories they are most interested in.

It is important to note that this was once only possible for editors to do; they were the ones who decided, based on their professional experience, which daily events would be highlighted or covered more in-depth.

This additional data on the behavior of information consumers allows for a level of customization that was previously impossible to achieve from more general metrics such as circulation and audience, both from the analog environment and too diffuse to allow an individualized level of granularity.

3.2 Predicting Social Engagement

The prediction models for trends can be applied to the world of journalism, expanding on various solutions for monitoring social media platforms which are now used in several business sectors.

What machine learning can do is use a pre-existing data set (combined with monitoring reports) to evaluate and predict what level of future engagement a particular news item will have; in other words, whether that particular news item has a greater or lesser chance of generating more interactions with that media channel's audience.

Companies like NewsWhip[2] are primarily engaged in

developing ML-based initiatives to optimize journalism production processes. This company has various products intended for use by newsrooms such as the ability to evaluate beforehand the potential for a tweet to go viral, thus giving journalists a heads up about a possible scoop.

This company uses "media predictive intelligence" and metrics such as "social velocity" to measure the rate of engagement with content on social media over time. It also uses regression algorithms to predict new interactions from the publication of a particular story, as in the example below.

In the article we use from *Estadao.com on* MPF recommendations to the government on the coronavirus, the NewsWhip tool for the Chrome browser allowed us to not only identify 644 *Facebook* interactions with the article (as any traditional monitoring algorithm would), but also to predict that it would have 262 new interactions within the next thirty minutes.

Figure 2 – Screenshot of NewsWhip solutions for predicting future engagement

Source: author's elaboration

This type of result, which can be combined with panels and viewed with other articles, offers a new layer of information for decision making in the newsroom.

DOI: 10.25200/BJR.v16n3.2020.1295

3.3 Allocation of Resources for Reporting

In addition to the data (including real time data) offered by analytics and records and historical series of interest captured by widely used tools like Google Trends[3], tools that use machine learning can even impact issues like logistics and the allocation of human or material resources in order that topics with greater growth potential are given more attention and space in engagement metrics.

An issue with rising potential can be passed on to a specialized team or sent to expert analysis, streamlining the normal process of generating content which will have repercussions and provide more in-depth coverage of immediate events.

Even resources like cars, motorcycles, mobile internet links, and helicopter flights can be optimized from predictive models that can cross over into the coverage routes of rising news and also traffic conditions, costs involved, and personnel available that day or on duty. The impact of this on the financial management of any news operation can be significant.

The allocation of the content and how it is structured on the pages that users will access can, in theory, also experience a significant increase in the level of personalization based on the combination of the various possibilities that we have previously listed.

Nowadays, based on the already widely explored logic of databases, most news portals generate a personalized page with only texts, photos and videos (for example, about a specific football team) and use simple techniques such as tags (labels), key words and categories.

Using machine learning solutions, this content can also be organized (without any great additional costs) according to a particular fan's preferences based on the data from that fan's previous interactions. For example, if a fan's profile shows a preference for watching videos of his or her favorite team, the page that will be presented will either highlight these items or have them at the top, so the user does not need to scroll down to find them, this way he or she has immediate access to the content they prefer.

Clearly this type of service is only available to subscribers who must log in to be identified, upload their usage profile, and then arrange the content and how it is presented according to their own personal tastes. This is a good example of what we call extended personalization, using digital subscription sales to monetize their

businesses, enhanced by the offer of different types of services and exclusive content, including virtual reality AI material or aggregated via augmented reality as described in Santos (2015).

4 Final considerations

This text discusses the strategic reconfiguration of data and its use in news organizations and how algorithms and artificial intelligence, specifically machine learning, are used as alternatives to increase the perceived value of the information product, a value which has reduced over time due to the fragmentation of audiences, the explosion of 'broadcasters', indirect competition from large technology companies, and cultural transformations of the digital media ecosystem.

This approach proposes to identify patterns of interest, to predict social engagement, and the allocation of resources for coverage as possible ways to expand the current level of personalization offered to news consumers.

The use of artificial intelligence and particularly machine learning is not perfect, nor is it free of problems or even risks.

For example, recent studies by Amadeu (2019) and Silva (2019) have drawn attention to situations where invasion of privacy, distortion of facts, and even racism and xenophobia were included in the results or propositions defined from ML algorithms.

As sequences of instructions, which only programmers have an in-depth knowledge of, artificial intelligence solutions and their algorithms, specifically machine learning, usually constitute non-transparent boxes, which are difficult for laymen to understand both in terms of the features offered as well as any internal risks or biases.

Inevitable and ubiquitous, code solutions are now part of the operational support of almost all contemporary human activities, representing both the beauty and magic of the first automata (Devaux, 1964) and the fictional or real risks and lack of control (Ellul, 1968) that so many authors, researchers, and screenwriters have already studied. Works by philosophers of technology like Feenberg (2002, 2010) try to balance out the pros and cons of a contemporary process.

The strategic use of data via algorithms is the business model that major global technology companies such as Google and *Facebook* use. They are partly responsible for reducing the perception

  DOI: 10.25200/BJR.v16n3.2020.1295

of the value of the traditional news product, including for its free use in their search tools and social media platforms, respectively.

As large data hubs, not only about news but also the social environment in which they operate, media companies were forced to take their content to the digital channels of information dissemination controlled by technology companies as a way to minimize the fragmented audiences that they helped create.

In this situation, the valorization of the available data becomes a path of strategic potential that must be considered. It is not only used in the aspects reported in this article: identifying patterns of interest; predicting social engagement; and allocating resources for coverage is much broader.

We believe that in order to face the central problem of society's reduced perception of the value of traditional news products, the effective use of algorithms, data, and emerging technologies and an increased level of personalization are paths which need to be explored.

The fight against fake news and the transitory centrality recovered due to the pandemic do not seem to be enough to resolve structural and previous issues.

This text is exploratory in nature and is far from establishing definitive answers for the complex network of possibilities and overlap between news making and technology.

NOTES

1 Data that can be accessed and used without any restrictions or costs, including commercially, and at most requires only that the source is credited.

2 www.newswhip.com/

3 https://trends.google.com.br/

REFERENCES

Amadeu, S. (2019). *Democracia e os códigos invisíveis*: *Como os algoritmos estão modulando comportamentos e escolhas políticas*. São Paulo: Edições Sesc (Coleção Democracia Digital).

Amaral, F. (2016). *Aprenda mineração de dados – Teoria e prática*. Rio de Janeiro: Alta Books.

Barbosa, S. (2007). *Jornalismo Digital em Base de Dados (JDBD) – Um paradigma para produtos jornalísticos digitais dinâmicos* [doctoral thesis, Universidade Federal da Bahia]. Repositório Institucional UFBA.

Barbosa, S. (2008). Modelo Jornalismo Digital em Base de Dados (JDBD) em Interação com a Convergência Jornalística. *Revista de la Sociedad Española de Periodística*, (*1)*, 87-106.

Barbosa, S. (2009). Modelo JDBD e o ciberjornalismo de quarta geração. In J. M. Flores Vivar, & F. E. Ramírez (Eds.), *Periodismo Web 2.0* (pp.271 – 283). Madrid, ES: Editorial Fragua.

Barbosa, S. (2011). Jornalismo em ambientes dinâmicos: Perspectivas, tendências e desafios para a criação de conteúdos em tempos de convergência. In A. L. Ureta, & K. M. Ayerdi (Eds.), *III Congreso Internacional de Ciberperiodismo y Web 2.0* (pp.7 – 17). Bilbao, Espanha: Universidad del País Vasco.

Barbosa, S., & Torres, V. (2013). O paradigma 'Jornalismo Digital em Base de Dados': Modos de narrar, formatos e visualização para conteúdos. *Galaxia*, (25), 152-164. DOI: 10.1590/S1982-25532013000200013

Carlson, M. (2014). The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms and journalistic authority. *Digital Journalism*, *3* (3), 416-431. DOI: 10.1080/21670811.2014.976412

Carreira, K. (2017). *Notícias Automatizadas – A evolução que levou o jornalismo a ser feito por não humanos* [master thesis, Universidade Metodista]. Biblioteca digital de teses e dissertações.

Clerwall, C. (2014). Enter the Robot Journalist: User's perception of automated content. *Journalism Practice*, 8 (5), 519 – 531. DOI: 10.1080/17512786.2014.883116

Coddington, M. (2014). Clarifying Journalism's Quantitative Turn. A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3 (3), 331–348. DOI: 10.1080/21670811.2014.976400

Coppin, B. (2010). *Inteligência artificial*. Rio de Janeiro: LTC.

Dalen, A. (2012). The Algorithms Behind the Headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice,* 6(5-6), 648-658. DOI: 10.1080/17512786.2012.667268

Devaux, P. (1964). *Autómatos, automatismo e automatização*. Lisboa: Editorial Gleba.

  DOI: 10.25200/BJR.v16n3.2020.1295

Diakopoulos, N. (2011, abril). *A Functional Roadmap for Innovation in Computational Journalism*. Nicholas Diakopoulos – Musings on Media. Retrieved from www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/

Ellul, J. (1968). *A técnica e o desafio do século*. Rio de Janeiro: Paz e Terra.

Feenberg, A. (2002). *Transforming technology*: *A critical theory revisited*. New York: Oxford University Press.

Feenberg, A. (2010). *Between reason and experience*. *Essays in technology and modernity*. Cambridge, MA: Mit Press.

Fidalgo, A. (2007). A resolução semântica no jornalismo online. In S. Barbosa (Ed.), *Jornalismo digital de terceira geração* (pp.93 – 102). Covilhã, PT: Livros LabCOM.

Fidalgo, A. (2004). Do poliedro à esfera: os campos de classificação. A resolução semântica no jornalismo online. Proceedings from *II Encontro Nacional da SBPJor*. Salvador, BA.

Graefe, A. (2016, jan.). *Guide to Automated Journalism*. Tow Center for Digital Journalism and James S. and John L. Knight Foundation. Retrieved from www.gitbook.com/book/towcenter/guide-to-automated-journalism/details

Grus, J. (2016). *Data Science do Zero*. Rio de Janeiro: Alta Books.

Lage, N. (1997). O lead clássico como base para a automação do discurso informativo. Proceedings from *20 Congresso Brasileiro de Pesquisadores da Comunicação Intercom.* Santos, SP.

Latar, N. (2015). The Robot Journalism in the Age of Social Physics: The end of human journalism? In G. Einav (Ed.), *The New World of Transitioned Media* (pp.65 – 80). Springer. DOI: 10.1007/978-3-319-09009-2_6

Lewis, S. & Usher, N. (2014). Collaboration and The Future of Journalism: A case study of the Hacks/Hackers global network. *Digital Journalism*, 2(3), 383–393. DOI: 10.1080/21670811.2014.895504

Machado, E. (2006). *O Jornalismo Digital em Base de Dados*. Florianópolis, SC: Calandra.

Maier, S. R. (2002). Numbers in the News: A Mathematics Audit of a Daily Newspaper. *Journalism Studies, 3* (4), 507-519. DOI: 10.1080/1461670022000019191

Meyer, P. (1973). *Precision Journalism. A reporter's introduction to social science methods*. Bloomington: Indiana University Press.

Nechushtai, E., & Lewis, S. (2019). What kind of news gatekeepers do we want machines to be? filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in human behavior,* 298–307. DOI: 10.1016/J.CHB.2018.07.043

Rubio, L., & Ruiz, M. (2020). Percepción de docentes universitarios, estudiantes, responsables de innovación y periodistas sobre el uso de inteligencia artificial en periodismo. *El profesional de la información*, 29 (1), 1–14. DOI: 10.3145/EPI.2020.ENE.09

Sánchez, J., & Ruiz, M. (2020). Inteligencia artificial y periodismo: una herramienta contra la desinformación. *Revista Cidob d'afers internacionals*, (124), 49–72. DOI: 10.24241/RCAI.2020.124.1.49

Santos, M. (2014). Textos Gerados por Software – Surge Um Novo Gênero Jornalístico? Proceedings from *XXXVII Congresso Brasileiro de Ciências da Comunicação*. Foz do Iguaçu, Paraná. Retrieved from www.intercom.org.br/papers/nacionais/2014/resumos/R9-2066-1.pdf

Santos, M. (2015). Jornalismo, mobilidade e realidade aumentada: notas sobre possibilidades de utilização. In J. Canavilhas, & I. Satuf (Eds.), *Jornalismo para Dispositivos Móveis: Produção, distribuição e consumo* (pp.103 –126). Covilhã, Portugal: UBI.

Santos, M. (2016a). Narrativas Automatizadas e a Geração de Textos Jornalísticos: A estrutura de organização do lead traduzida em código. *Brazilian Journalism Research*, 12(1), 160-185. DOI: 10.25200/BJR.v12n1.2016.757

Santos, M. (2016b). Internet das Coisas e Sistemas Inteligentes no Jornalismo. *Comunicação e Inovação*, 17(34), 21-39. DOI: 10.13037/ci.vol17n34.3769

Santos, M. (2016c). Jornalismo e Internet Das Coisas – Notas sobre tipologia e modelos de utilização. *Revista Lationoamericana de Ciencias de La Comunicación*, 12 (22), 134–145. Retrieved from www.alaic.org/revista/index.php/alaic/article/view/621/380

Santos, M. (2019a). Narrativas e desenvolvimento de conteúdo imersivo: Aplicação de sistemas bi e multissensoriais de realidade virtual no jornalismo. *Revista da Intercom*, 42 (3), 133-150. DOI: 10.1590/1809-5844201937

Santos, M. (2019b. A datificação de um campo de conhecimento. *Organicom USP*, (31), 145–157. Retrieved from www.revistas.usp.br/organicom/article/view/161444/159283

Saquete, E., Tomás, D., Moreda, P., Martínez-Barco, P., & Palomar, M. (2020). Fighting post-truth using natural language processing: a review and open challenges. *Expert systems with applications*. DOI:

  doi: 10.25200/bjr.v16n3.2020.1295

10.1016/J.ESWA.2019.112943

Silva, T. (2019). Linha do tempo do racismo algorítmico. *Blog do Tarcízio Silva*. Retrieved from https://tarciziosilva.com.br/blog/posts/racismo-algoritmico-linha-do-tempo

López, J. M., Bran, C., & Requeijo, S. (2018). Uso de bots y algoritmos para automatizar la redacción de noticias: percepción y actitudes de los periodistas en españa. *El profesional de la información*, 27(4), 750–758. DOI: 10.3145/EPI.2018.JUL.04

Višnovský, J., Ungerová, M., & Kubíková, K. (2019). Robo-journalism and its implementation in editorial practice. *Proceedings of the NORDSCI Conference*. Retrieved from https://files.eric.ed.gov/fulltext/ED603467.pdf

MÁRCIO CARNEIRO DOS SANTOS. PhD from the Intelligence Technologies and Digital Design (TIDD) program at PUC–SP. Full professor of the Post-Graduate Program in Design and Post-Graduate Program in Communication – Professional Master's Degree, both at UFMA. Professor at the Department of Social Communication in the field of Digital Network Journalism. Coordinator of LABCOM (Media Convergence Laboratory). Leader of CNPq's Technology and Digital Narratives research group (TECND). Email: mcszen@gmail.com

TRANSLATED BY: LEE SHARP